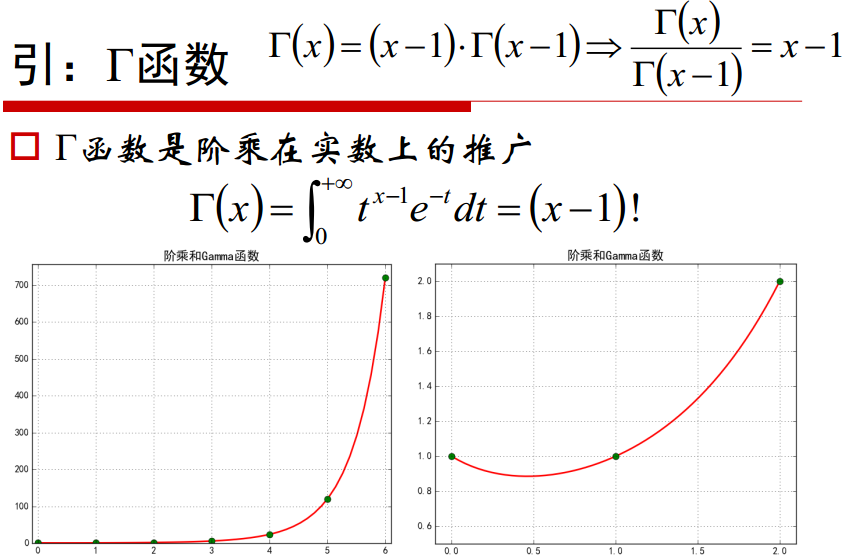
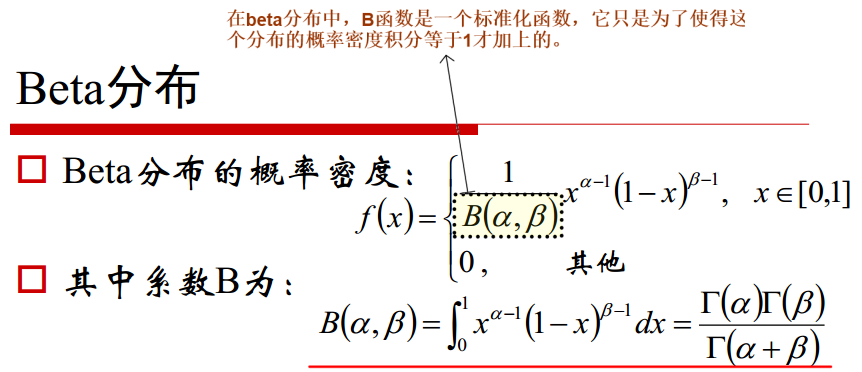
1. **Gamma函数**

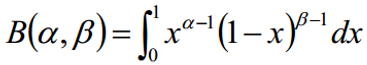


1. **Beta分布**

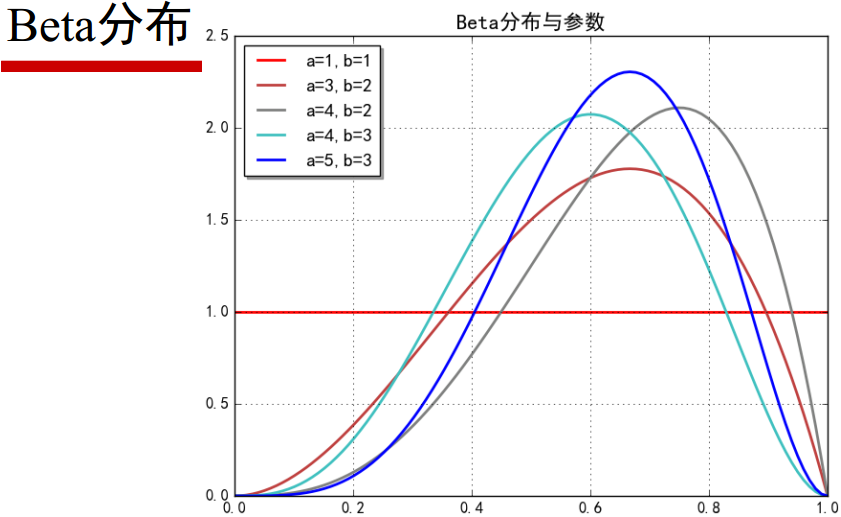


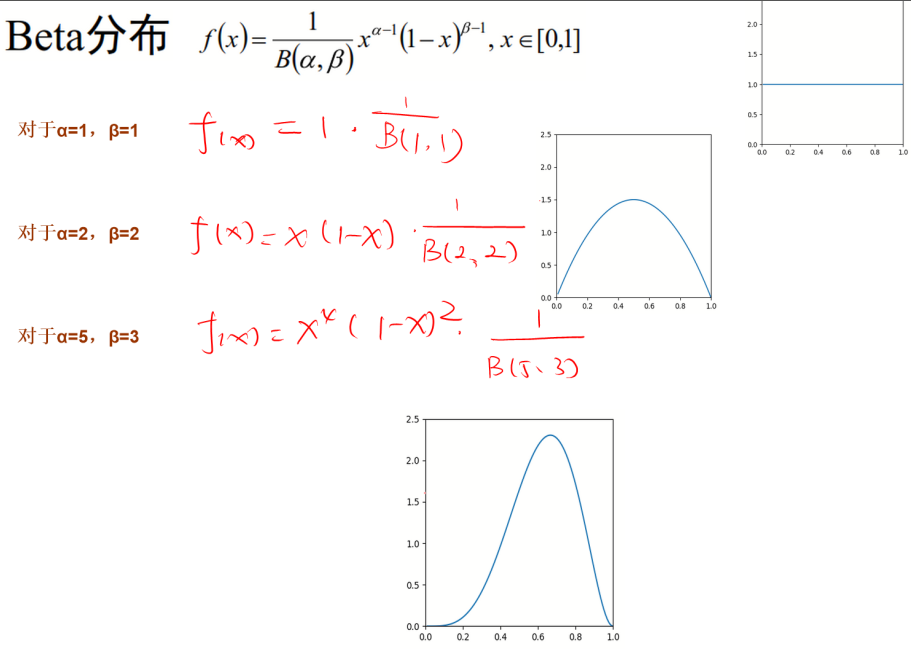
**因为beta分布概率密度在[0,1]上积分等于1**

**所以**

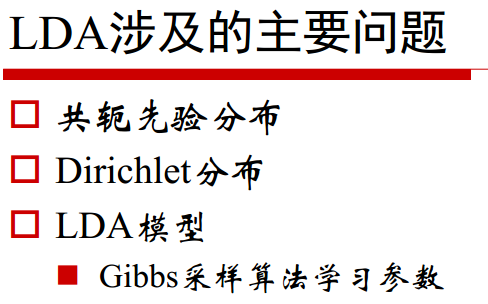
**其中为常数，故**

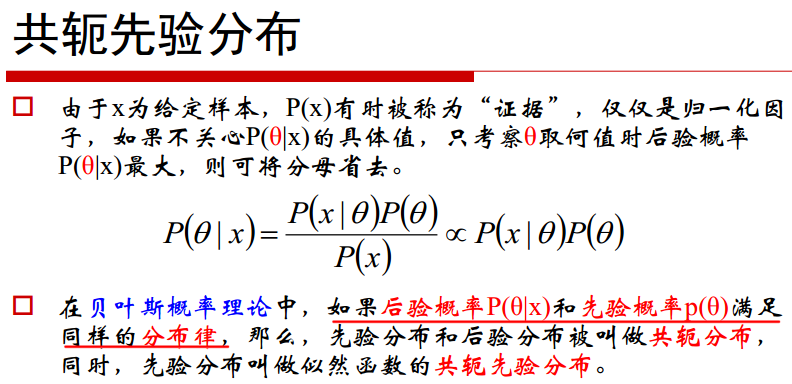


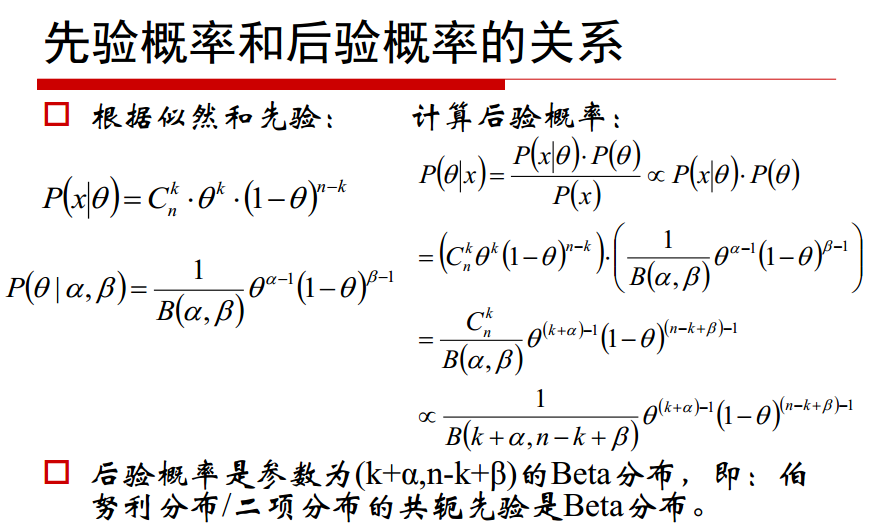




1. **LDA涉及的问题**

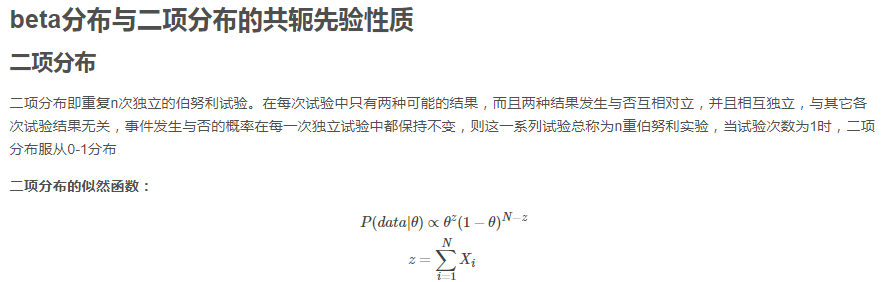


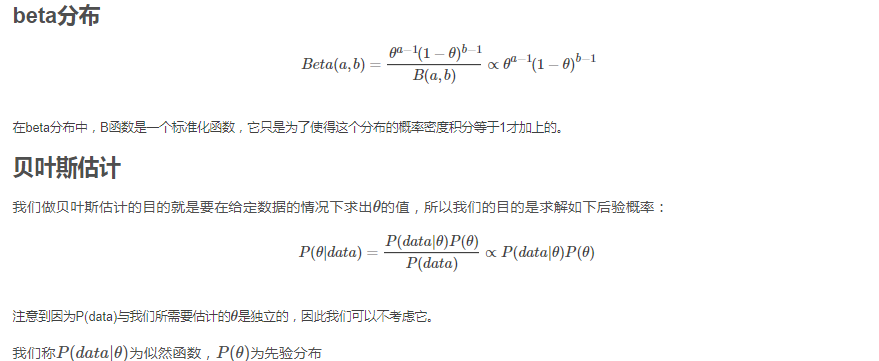


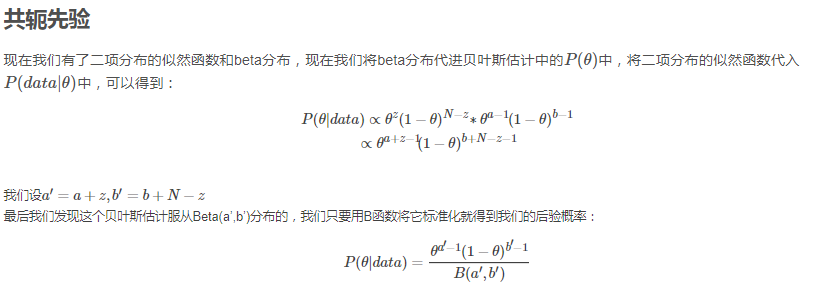




1. **Beta分布与二项分布**

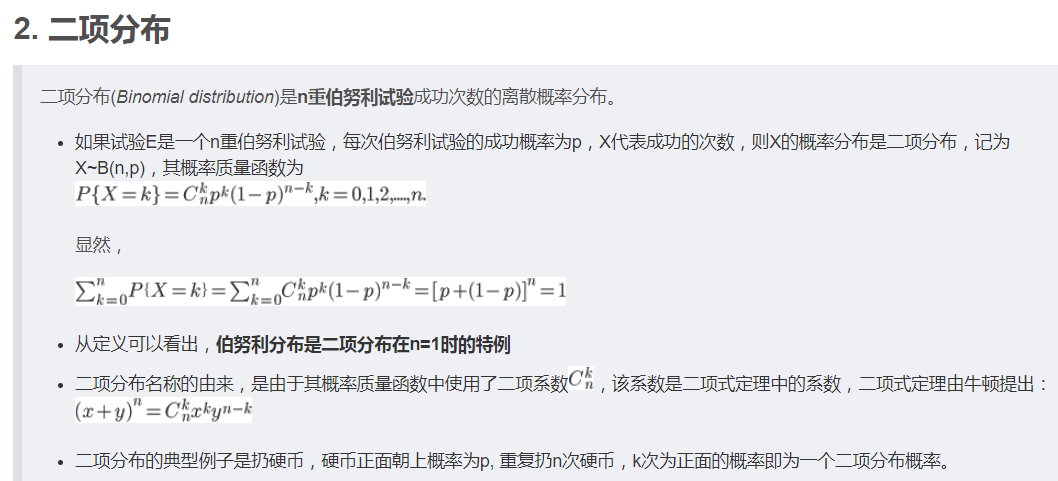






1. **从二项分布到多项式分布**

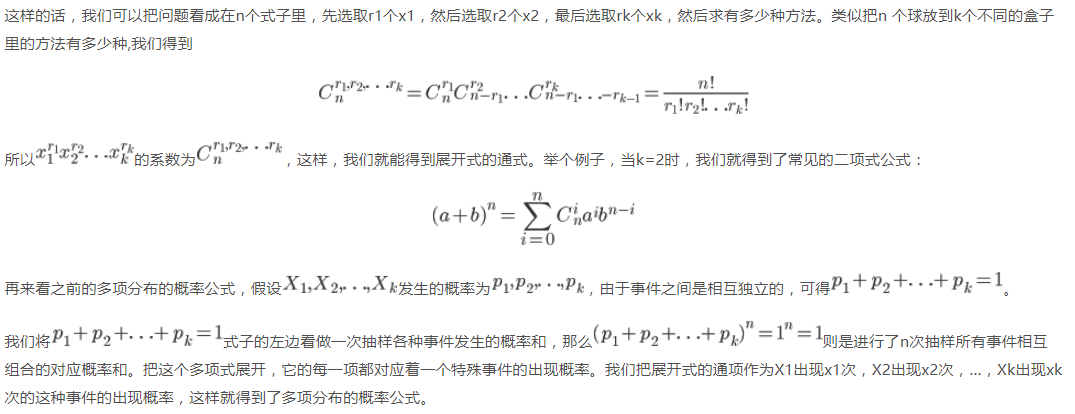
**二项分布**





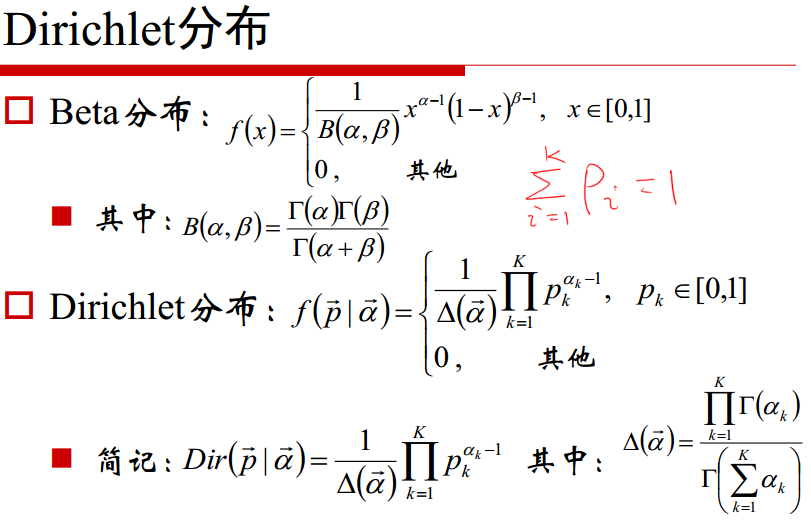
**多项式分布**

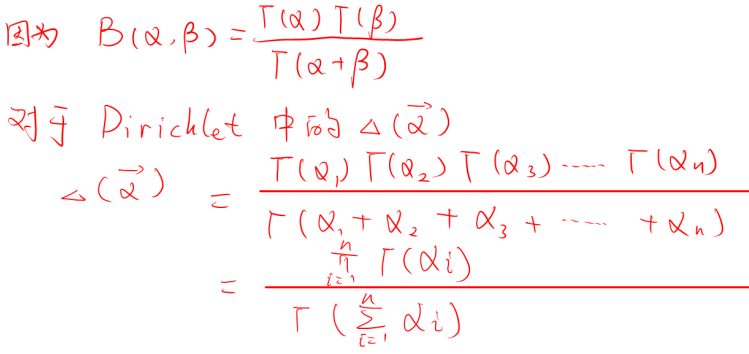


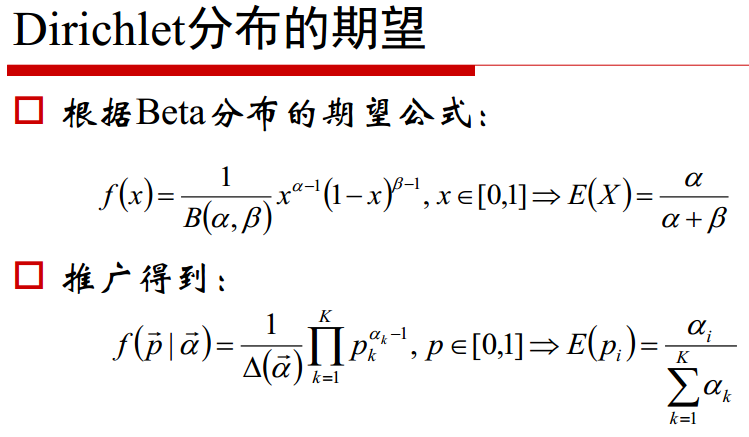


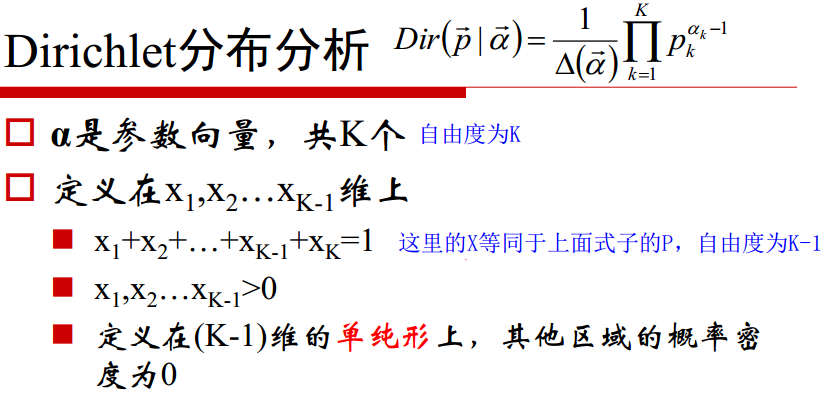


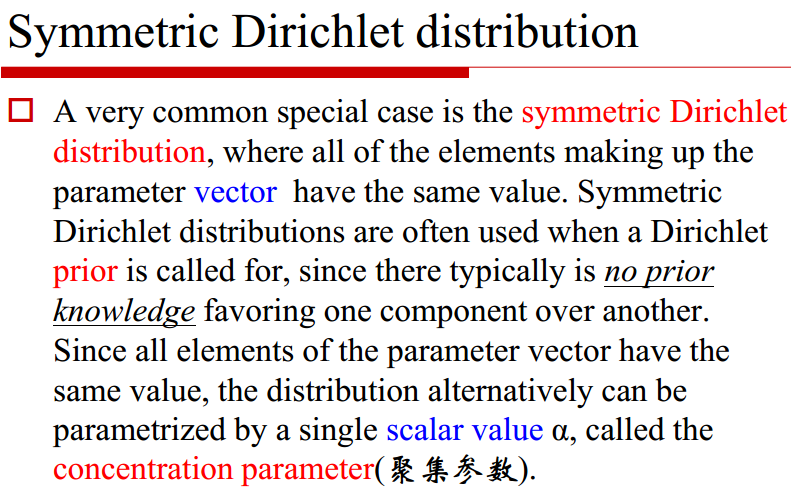
1. **从beta分布到dirichlet分布**



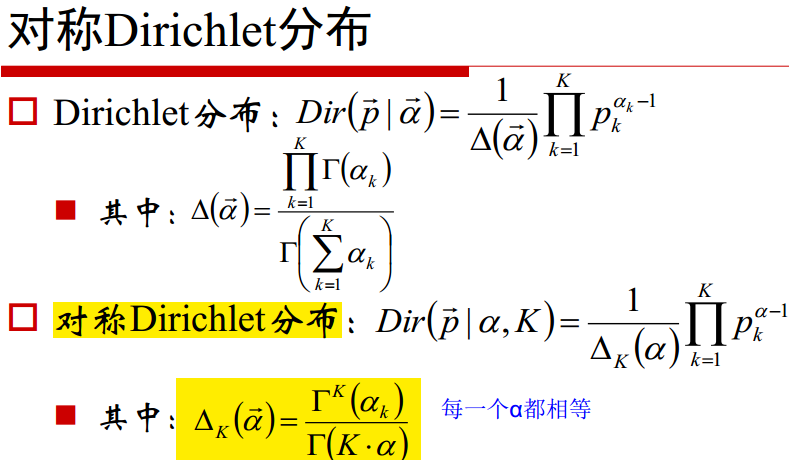


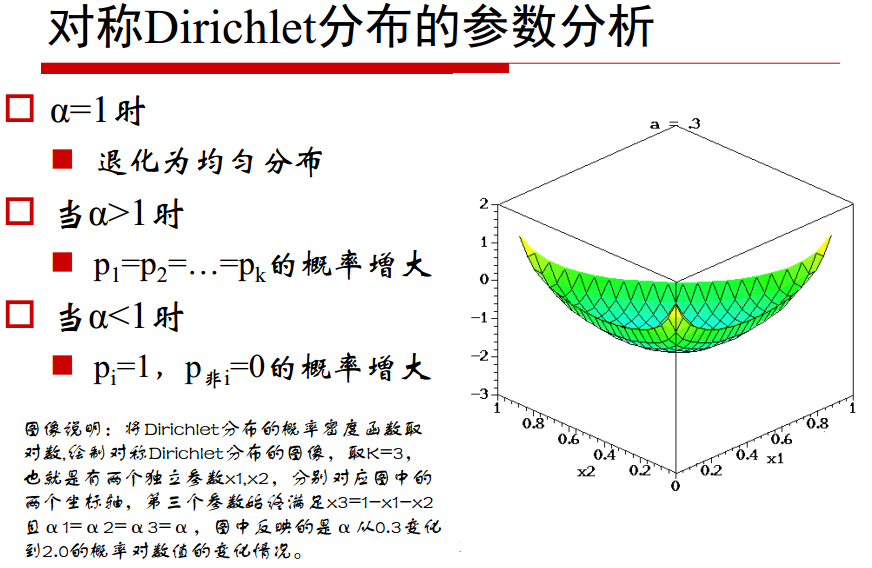


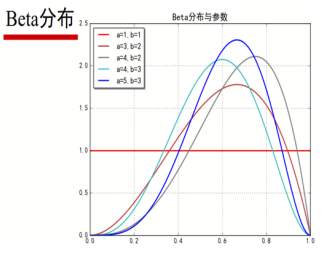
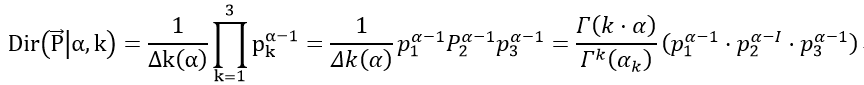


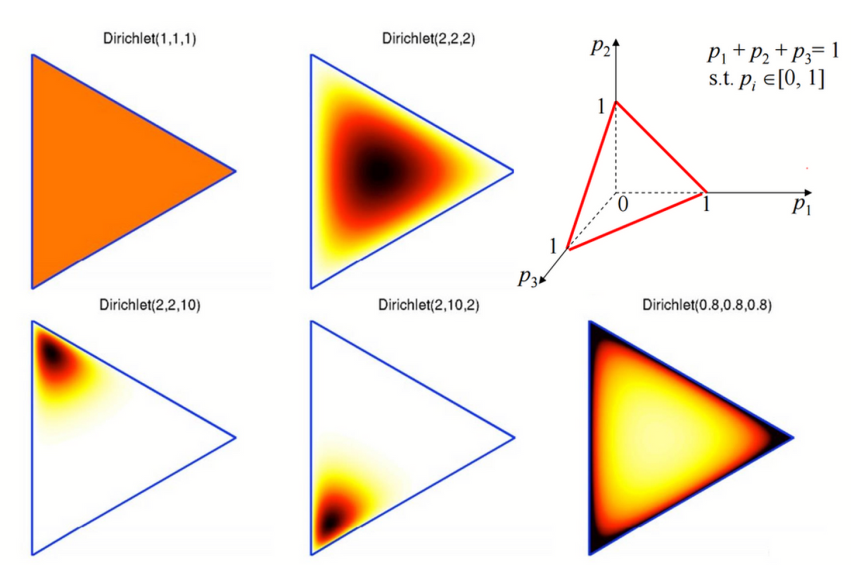


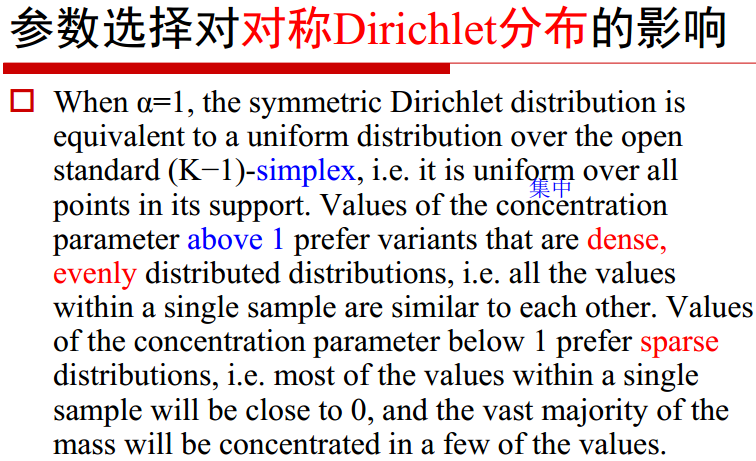
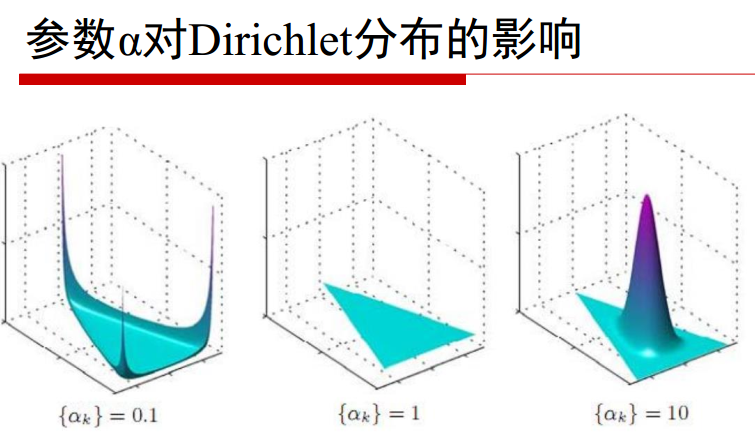




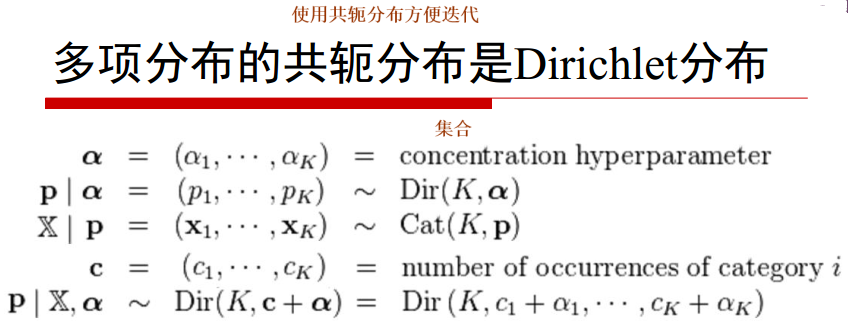










1. **LDA的解释**

每篇文章 ( 长度为Nm ) 都有各自的主题分布，主题分布是多项分布，该多项分布的参数，服从Dirichlet分布，该Dirichlet分布的参数为α；（有多少个主题就有多少个α，自由度为主题个数）

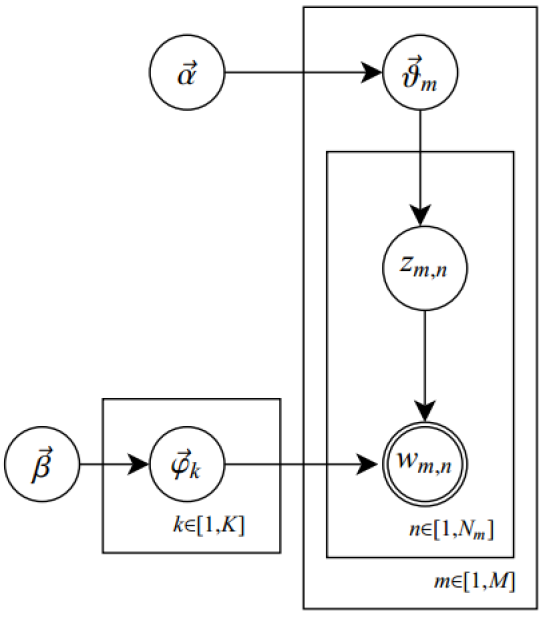
这是主题分布θ，共有m篇文章，一共涉及了K个主题；自由度为K-1，因为.

从主题分布中拿出一个主题Z

每个主题都有各自的词分布，词分布为多项分布，该多项分布的参数服从Dirichlet分布，该Dirichlet分布的参数为β；（有多少个词就有多少个β，自由度为词个数）

这是词分布θ，一共涉及了n个词；自由度为n-1，因为.

在对应主题采出对应的词



  1. 对每一篇文档，从主题分布中抽取一个主题；

  2. 从上述被抽到的主题所对应的单词分布中抽取一个单词；

  3. 重复上述过程直至遍历文档中的每一个单词。

